

Best Practice Guide for Selecting a Pre-Employment Assessment

Choosing assessments to use in an employment selection context involves consideration of technical, legal, operational, and other issues. When companies decide to explore the use of assessments for employment selection employers are faced with a virtually endless selection of assessments and providers. The following critical issues should be considered when choosing employment assessments. Additionally, assessments should be developed and maintained under current and specific industry guidelines. Among these are:

- American Psychological Association (1999), Standards for educational and psychological Testing
<http://www.apa.org/science/standards.html>
- Department of Labor (2000), Testing and assessment: An employer's guide to good practices
http://www.onetcenter.org/dl_files/empTestAsse.pdf
- Equal Employment Opportunity Commission (1978), Uniform guidelines on employee selection procedures
http://www.dol.gov/dol/allcfr/Title_41/Part_60-3/toc.htm
- Society for Industrial and Organizational Psychology (2001), Principles for the validation and use of personnel selection procedures
http://www.siop.org/_Principles/principles.pdf

The critical issues below describe those considerations that are common and critical to nearly all employment assessment situations. Some additional issues that are more situational dependent include issues such as language availability, delivery media, customization, integration with ATS or HRIS systems, etc. An Assessment Evaluation Worksheet is included at the end of this document for your reference.

1. Characteristics measured by the assessment should map to characteristics required for successful job performance.

When an assessment is used in an employment setting, there is a need to demonstrate that assessment scores are related to job performance. This relationship should first be demonstrated by conceptually mapping the content of the assessment (typically described in the assessment manual) to the content of the job (typically described by a job description or competency model).

For example, the Raven's Advanced Progressive Matrices is a measure of cognitive ability. To demonstrate that the Raven's content is conceptually linked to specific job content, it would be necessary to identify competencies necessary for success on the target job that involve cognitive ability (e.g., Analytical Thinking).

2. The assessment should have adequate reliability; that is, the assessment should yield consistent results.

Perfect reliability would be said to exist if an assessment taker achieved the same scores on an assessment taken repeatedly. On the other hand, low reliability would exist if a candidate's scores were the result of pure chance (e.g., a coin flip), and therefore differed substantially over many assessments.

Reliability is typically reported as a number ranging between 0 and 1, with reliabilities of .70 and higher conventionally considered as adequate. For example, according to the Department of Labor (1999), reliabilities:

- less than .70 "may have limited applicability,"
- between .70 and .79 are "adequate,"
- between .80 and .89 are "good," and
- above .90 are "excellent."

3. Results of the assessment should be useful for making the inference(s) that the assessment user desires to make.

Scores on employment assessments are used to predict or infer a candidate's job potential. Therefore, it is important that information be available showing that assessment scores are statistically related to job performance in jobs similar to the one(s) that the assessment user plans to use the assessment for. The statistical information should show that people with higher assessment scores tend to do better on the job than people with lower assessment scores.

This type of statistical validity evidence, known as criterion-related validity, is typically reported as a number ranging between 0 and 1. According to the Department of Labor (1999), validities:

- above .35 are "very beneficial,"
- between .21 and .35 are "likely to be useful,"
- between .11 and .20 "depends on circumstances," and
- below .11 are "unlikely to be useful."

4. The assessment should be unbiased toward protected subgroups.

Assessment providers should provide evidence that the assessment does not contain bias on the basis of race, sex, or age. What this means is that if one race, sex, or age group obtains lower average assessment scores than members of another group, the differences in assessment scores should also be reflected in the differences in the job performance of members of the groups (Department of Labor, 2000).

Absence of bias does not mean that the assessment will have similar results for different groups of people. For example, in an employment decision context, more women than men may score high on a assessment of verbal ability, and more men than women may score high on a assessment of mechanical ability. In these situations, the assessments would not be considered biased if women and men with similar assessment scores achieved similar job performance.

The assessment manual and independent reviews, such as those found in the Buros Mental Measurements Yearbook, available online, should provide information that helps to evaluate whether or not an assessment is unbiased. If this information is not readily available, the assessment provider should be able to provide statistical evidence or references related to the assessment's fairness and lack of bias.

5. Information must be available to determine how high or low an assessment score is.

Normative data, or "norms," should be available from the assessment provider to provide a benchmark for interpreting an individual's scores relative to other candidates who completed the assessment. Ideally, norms are available for a group of people that are in the position for which testing is being used (e.g., for candidates being tested for managerial positions, it is ideal to have managerial norms available). Such norms provide more precision in the interpretation of candidates' scores.

For information on selecting a norm group, please see our Best Practice Guide for Choosing Norm Groups in Employment Selection whitepaper <http://pearsonassess.com/haiweb/Cultures/en-US/Pearson/Community/TalentAssessment/Research/BestPractices/reslist.htm>.

6. Administration and scoring mode should align with available resources.

Computer-based administration formats have several advantageous features including consistency in administration, automatic scoring, and real time delivery of results in an interpretive report format. Paper-and-pencil tests, on the other hand, are useful when computer resources are not available, or if the number of test administrations at a given time is likely to exceed the number of computers available. Paper-and-pencil tests are typically scored manually.

Test Evaluation Worksheet

**Characteristics measured by the test
(e.g., critical thinking, mathematical ability):**

**Job content (e.g., competencies, responsibilities,
or behaviors) that the characteristics measured by
the assessment link to (e.g., the need to make sound
decisions on the fly; the need to consider issues from
multiple perspectives):**

Target job to be selected for (e.g., mid-level managers):

Norms available:

Reliability:

Validity (criterion-related):

Assessment Bias:

Administration time:

Administration and scoring mode:

Ease of working with the assessment provider:

Potential Additional Issues

- Type of report that's produced with online assessments (e.g., is the information provided clear and actionable?):
- Languages available:
- Ease of integrating assessment results with applicant tracking and HRIS systems:
- Suitability of test results for use in professional development:
- Other:

Overall Evaluation: